Tools and Techniques for Automating the Analysis of EOSDIS Data

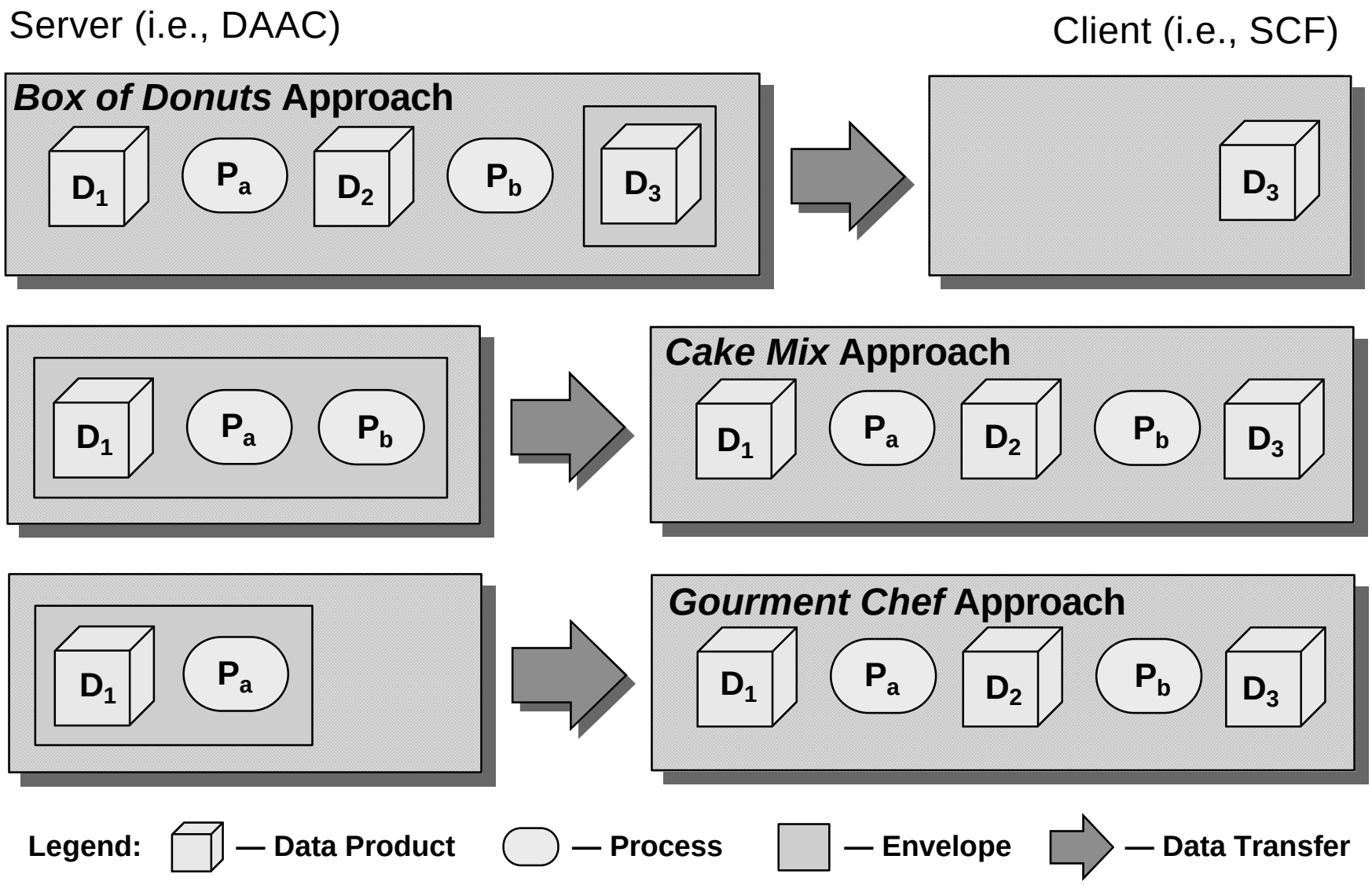
- **Principal Investigator:**
  - **William Emery, U. of Colorado — Aerospace Engineering**
- **Co-Principal Investigator:**
  - **Randal Davis, U. of Colorado — LASP**
- **Co-Investigators:**
  - **Frank LoPinto, Computer Sciences Corporation**
  - **James Maslanik, U. of Colorado — Aerospace Engineering**
  - **Lewis Pinson, U. of Colorado, Colorado Springs — Computer Science**
  - **Louis Reich, Computer Sciences Corporation**
  - **Carol Wessman, U. of Colorado — CIRES**
  - **Richard Wiener, U. of Colorado, Colorado Springs — Computer Science**
- **Chief Developers:**
  - **Ann Bateson, U. of Colorado — CIRES**
  - **Bob Biro, U. of Colorado — CCAR**
  - **Rima Dolgonos, U. of Colorado — LASP**
  - **David Judd, U. of Colorado — LASP**

Objectives

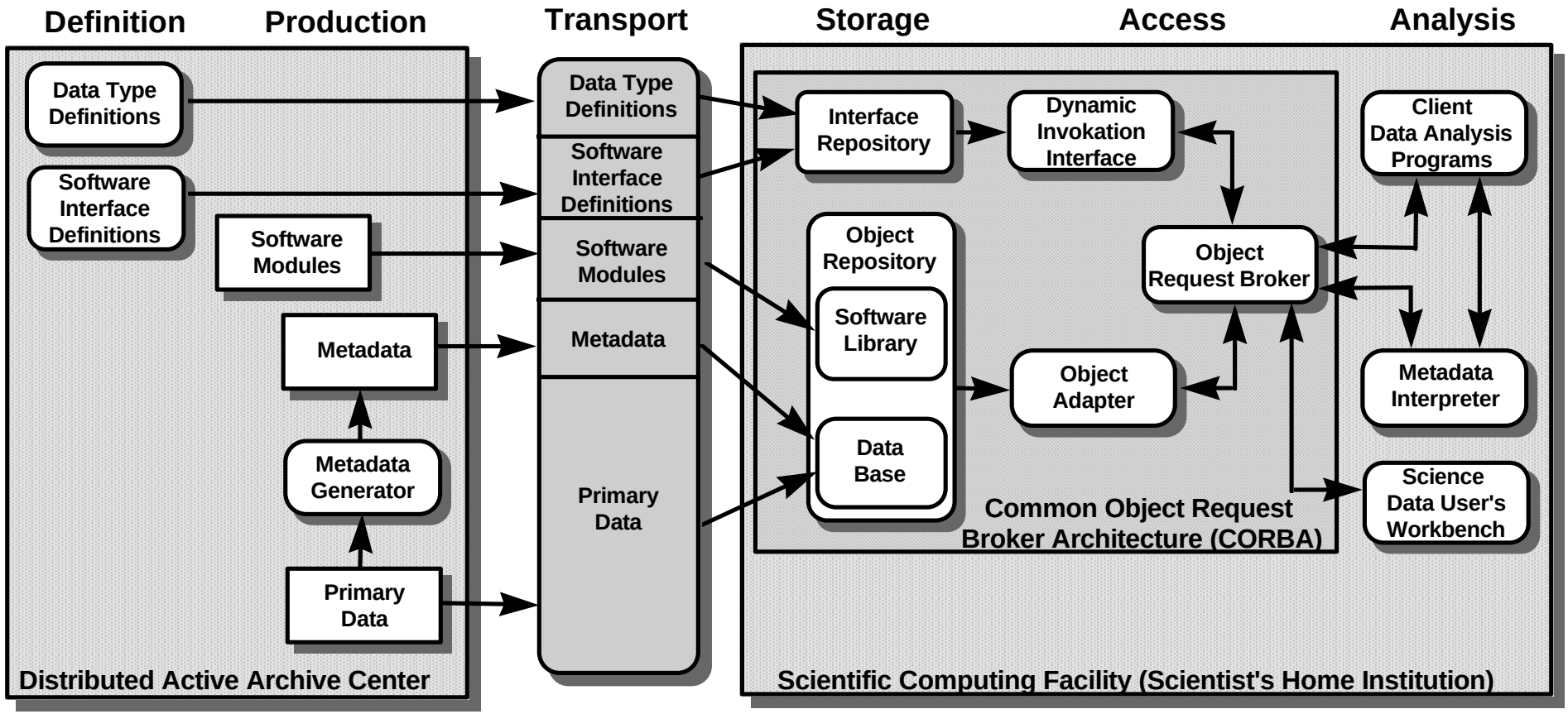
1. **Make it easier for scientists who are not expert in a particular EOSDIS dataset to use that dataset in their analysis**
2. **Make it possible for scientists who are experts in particular datasets to analyze much larger volumes of data**

- **To do this, we are:**
  - **Providing a framework for generating custom data products**
  - **Allowing users to tap into the large base of existing data analysis tools**
  - **Helping users to set up those tools for use in their research**
  - **Connecting those tools to a scientist's own data analysis software**
  - **Adding intelligent software agents to the analysis process**

Three Models of Distributed/Distributable Data Systems



Our Initial Design

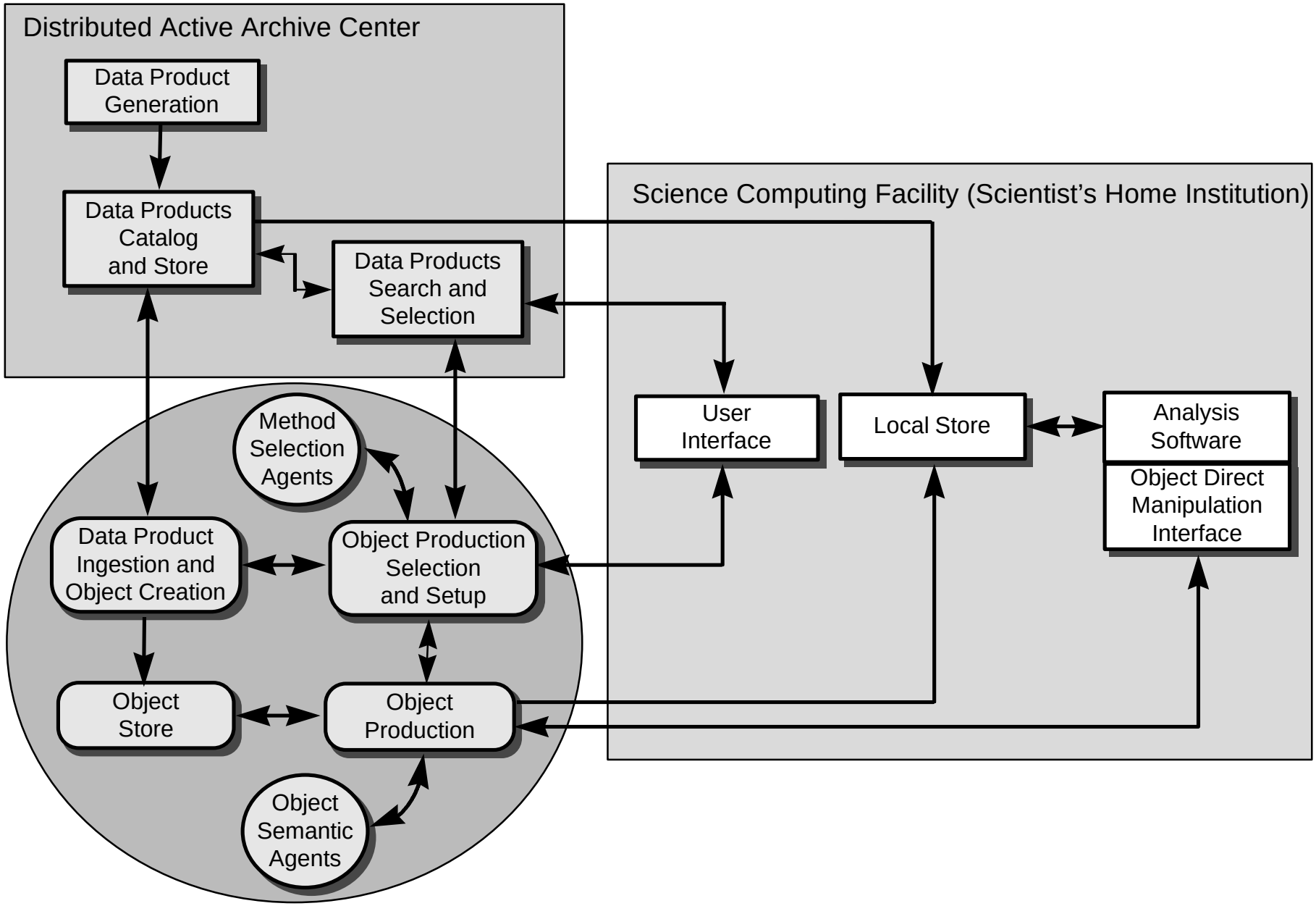


Components To Be Developed Are Indicated as Rounded Rectangles

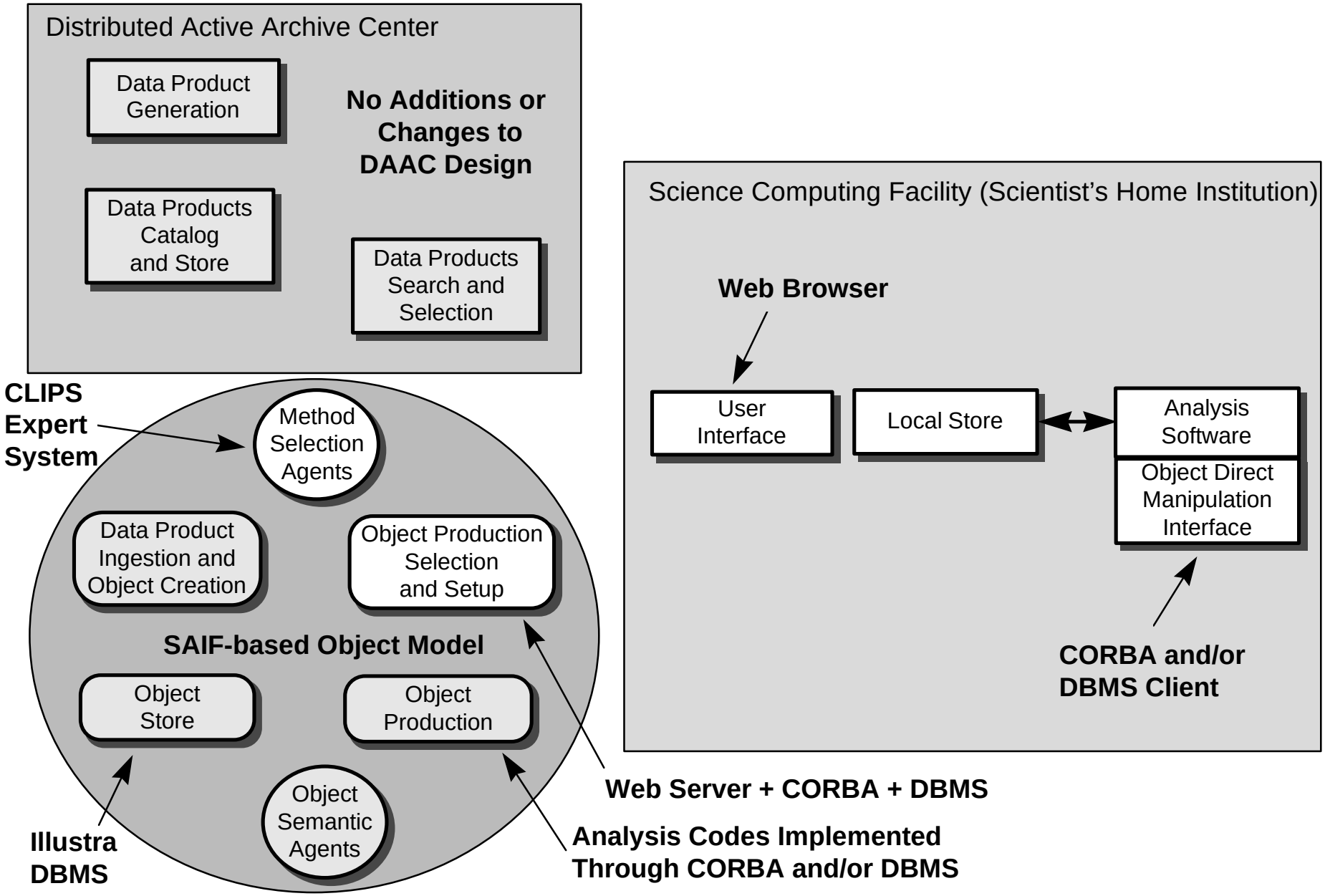
Accomplishments to Date

- Analyzed several existing object models to see how they would work in real science applications
  - Decided to use SAIF as the model for our initial prototype
- Extended SAIF with a set of objects for AVHRR data processing
- Selected and installed DEC Objectbroker as our CORBA implementation
- Adapted existing software for AVHRR processing to work with CORBA
- Installed Illustra for use as an object store
- Developed a World Wide Web interface for users to select data objects they want and to set up the process for generating those objects
- Combined all of the above into a simple prototype
- Tested the prototype
- Demonstrated that the basic approach is sound . . .
  - For example, CORBA is appropriate for use in scientific applications
- . . . but there were problems with the initial design
  - Biggest problem is that legacy code is just too tough to port easily
  - CORBA complexity and cost is also a big concern
- Revamped the design and now implementing it

Updated Design



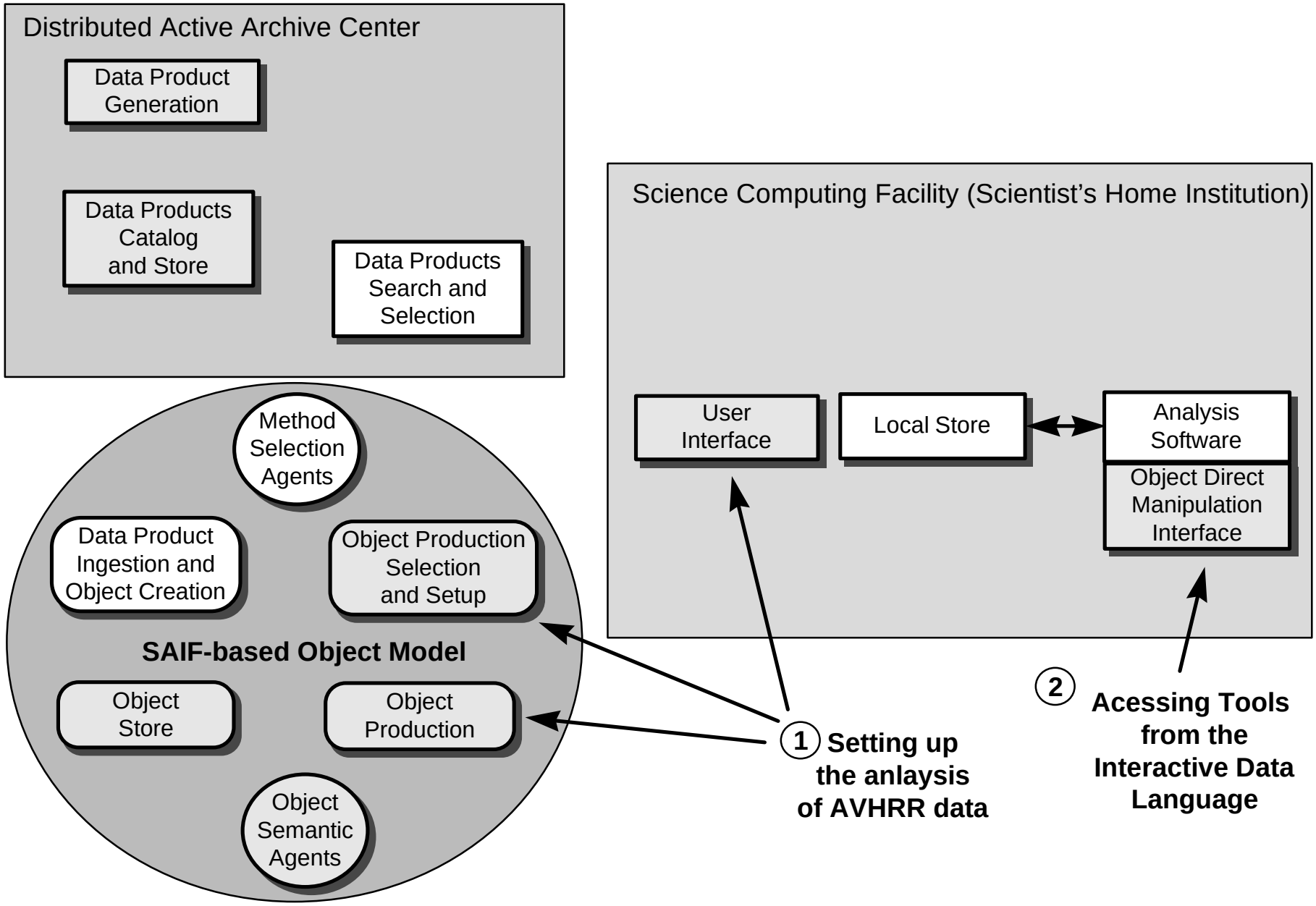
Implementation



Key Future Activities

- **Add more software for analyzing AVHRR data, such as tools for snow cover and fire detection**
- **Extend to other data sets by supporting specific products for SSM/I (water vapor, precipation, etc.) and Landsat TM (spectral unmixing)**
- **Implement the Object Production Selection and Setup function in a generic fashion, so that it will work with any data set and user application**
  - **Add in prototype Method Selection Agents to provide assistance to users in selecting and configuring methods for a specific analysis task**
- **Determine how to implement the Data Product Ingest and Object Creation function and its associated interface to the DAAC**
  - **Allow the user to select a set of data products for immediate analysis and also to specify a standing order for recurring processing**
- **Develop and demonstrate use of intelligent agents for making data processing decisions based on data semantics**
- **Provide access to analysis tools from the Interactive Data Language and other widely used data analysis systems**

What We Are Demonstrating Today



Sample Processing Path For AVHRR Data